RGe ISSN: 2177-3246

MODELING THE DASHBOARD PROVENANCE

MODELAGEM DA PROVENIÊNCIA DO PAINEL

MODELADO DE LA PROCEDENCIA DEL PANEL

10.56238/revgeov17n4-075

Johne Jarske, Jorge Rady, Lucia V. L. Filgueiras, Leandro M. Velloso, Tania L. Santos

ABSTRACT
Organizations of all kinds, whether public or private, profit-driven or non-profit, and across various industries and sectors, rely on dashboards for effective data visualization. However, the reliability and efficacy of these dashboards rely on the quality of the visuals and data they present. Studies show that fewer than a quarter of dashboards provide information about their sources, which is just one of the expected pieces of metadata when provenance is seriously considered. Provenance is a record that describes people, organizations, entities, and activities that had a role in the production, influence, or delivery of a piece of data or an object. This paper aims to provide a provenance representation model that enables standardization, modeling, generation, capture, and visualization, specifically designed for dashboards and their visual and data components. The proposed model will provide a comprehensive set of provenance metadata, enabling users to evaluate the quality, consistency, and reliability of the information presented on dashboards. This will provide a clear, precise understanding of the context in which a specific dashboard was developed, ultimately leading to better decision-making.

Keywords: Provenance Model. Dashboard. Visualization. User Experience.

RESUMO
Organizações de todos os tipos, sejam públicas ou privadas, com fins lucrativos ou sem fins lucrativos, e de diversos setores e indústrias, dependem de painéis (dashboards) para uma visualização eficaz de dados. No entanto, a confiabilidade e a eficácia desses painéis dependem da qualidade dos elementos visuais e dos dados que apresentam. Estudos mostram que menos de um quarto dos dashboards fornece informações sobre suas fontes, o que representa apenas uma das peças de metadados esperadas quando a proveniência é considerada de forma adequada. Proveniência é um registro que descreve pessoas, organizações, entidades e atividades que tiveram papel na produção, influência ou entrega de um dado ou objeto. Este artigo tem como objetivo fornecer um modelo de representação de proveniência que permita a padronização, modelagem, geração, captura e visualização, especificamente projetado para dashboards e seus componentes visuais e de dados. O modelo proposto fornecerá um conjunto abrangente de metadados de proveniência, permitindo aos usuários avaliar a qualidade, consistência e confiabilidade das informações apresentadas nos dashboards. Isso proporcionará uma compreensão clara e precisa do contexto no qual um determinado painel foi desenvolvido, resultando, em última análise, em melhores tomadas de decisão.

Palavras-chave: Modelo de Proveniência. Dashboard. Visualização. Experiência do Usuário.

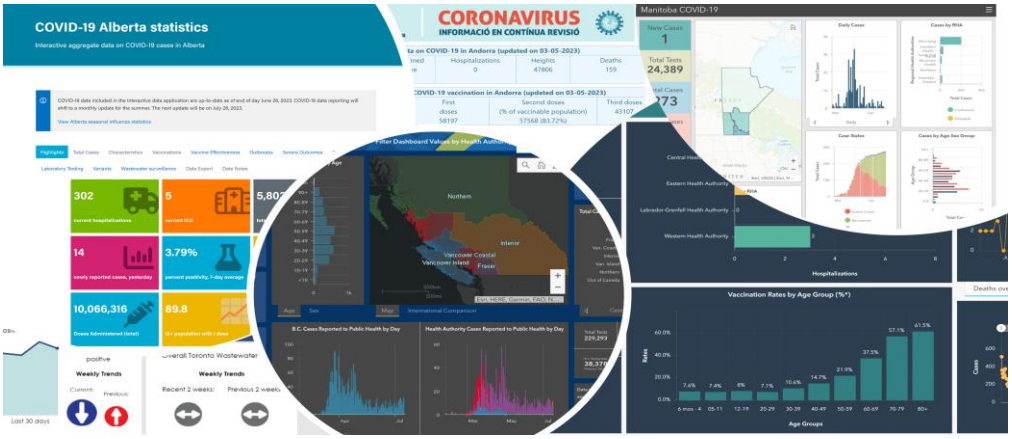
RESUMEN
Las organizaciones de todo tipo, ya sean públicas o privadas, con fines de lucro o sin fines de lucro, y de diversos sectores e industrias, dependen de los paneles de control (dashboards) para una visualización eficaz de datos. Sin embargo, la confiabilidad y la eficacia de estos paneles dependen de la calidad de los elementos visuales y de los datos que presentan. Los estudios muestran que menos de una cuarta parte de los dashboards proporciona información sobre sus fuentes, lo cual representa solo uno de los metadatos esperados cuando la procedencia se considera de manera adecuada. La procedencia es un registro que describe a las personas, organizaciones, entidades y actividades que tuvieron un papel en la producción, influencia o entrega de un dato u objeto. Este artículo tiene como objetivo proporcionar un modelo de representación de procedencia que permita la estandarización, modelado, generación, captura y visualización, específicamente diseñado para dashboards y sus componentes visuales y de datos. El modelo propuesto proporcionará un conjunto integral de metadatos de procedencia, permitiendo a los usuarios evaluar la calidad, consistencia y confiabilidad de la información presentada en los dashboards. Esto proporcionará una comprensión clara y precisa del contexto en el que se desarrolló un determinado panel, lo que en última instancia conducirá a una mejor toma de decisiones.

Palabras clave: Modelo de Procedencia. Dashboard. Visualización. Experiencia del Usuario.

1 INTRODUCTION

Figure 1

Web-based COVID-19 Dashboards



The use of dashboards for information dissemination and decision-making is widespread across almost all companies and public and private organizations, profit and non-profit, spanning various industry sectors. As highlighted by Sarikaya et al. [1], dashboards have become key tools for supporting data-driven decision-making.

Although dashboards are popular and useful for data visualization, their effectiveness and reliability depend on the quality of the data they present. The reliability of dashboards is closely related to the curation of the data, which involves providing clear information about the data's origin and the dashboard's visual components [2]. The research field that studies the origin of a digital or material object is known as provenance. According to Shankar et al. [2], provenance provides details on how a given resource is collected, transformed, normalized, and made available to the user. The Word Wide Web Consortium (W3C) published in 2013 the W3C-PROV, a standard for provenance that can be used to describe how data was gathered to ensure it is utilized in a meaningful way, ascertain who owns an object and their rights over it, evaluate information to decide if it is reliable, check if the steps taken to achieve a result are in line with given specifications, and replicate how something was created. The W3C defines provenance as a record that describes people, organizations, entities, and activities that had a role in the production, influence, or delivery of a piece of data or an object [3].

The emergence of the COVID-19 pandemic led universities, businesses, and governments worldwide to create dashboards to disseminate indicators such as the number of infected individuals, deaths, hospitalizations, and, later, vaccination metrics. According to Shankar et al. [2], on the one hand, dashboards such as those provided by the Centers for

Disease Control and Prevention [4] and Johns Hopkins University [5] provided detailed information on the provenance of the data. On the other hand, other dashboards were questioned due to a breakdown in the data provenance chain. The author cites the notable case involving the state of Florida's COVID-19 dashboard [6], where data scientist Rebekah Jones was allegedly fired for refusing to manipulate infection rate data to expedite economic reopening plans.

The Infovis for Public Health Research Project, conducted between 2020 and 2022, with funding from the Pan American Health Organization (OPAS), in response to the demand of the Department of Monitoring and Evaluation of the Brazilian Unified Health System - Executive Secretariat of the Ministry of Health (DEMAS/SE-MS - *Departamento de Monitoramento e Avaliação do SUS - Secretaria Executiva do Ministério da Saúde do Brasil*), evaluated the experience of public health managers with the dashboards provided by the Strategic Management Support Room (SAGE - *Sala de Apoio à Gestão Estratégica*). Relevant issues regarding the reliability of the information displayed on these dashboards were identified. During the interviews conducted by the project team, concerns were raised about the lack of information on the data source, the absence of identification of the responsible party for the information presented, and the need for clarity on the frequency and periodicity of data updates - common pieces of information in data provenance.

The term "provenance" originated in the fine arts [2], [7], [8], [9], [10], [11], [12], where it refers to information about the creation, ownership history, custody chain, modifications, or influences an artwork or manuscript has undergone.

In computing, the notion of provenance has been adopted and extended to apply to digital artifacts, particularly in data and user interaction, cognition, visualization, and visual analysis. In computing, provenance goes beyond origin or history to include the transformation process and other contextual information [2].

Provenance is inherently context-dependent, as different domains have unique requirements and use it in diverse ways. It is essential to understand the specific problems that provenance addresses within each context and tailor its implementation accordingly [11].

In an effort to create a comprehensive provenance model that accommodates diverse use cases, the World Wide Web Consortium (W3C) developed the W3C-PROV model in 2013. This model has become a de facto standard, with research and applications conducted in various domains [13]. The W3C-PROV model defines provenance as information about entities, activities, and people who produce a given thing or data. It can be used to evaluate its quality and reliability [14], [15].

Recent studies have highlighted the need for further investigation into provenance in

dashboards. Zhang et al. [16] highlight the credibility crisis that emerged amid the COVID-19 pandemic. On the one hand, the public seeks more reliable information. On the other hand, government agencies, health experts, and dashboard designers express reservations about the reliability of their multiple sources, which could lead to confusion and even undermine trust in the authorities. Shankar et al. [2] highlighted the issue of trustworthiness surrounding certain COVID-19 dashboards, which comes into question when the provenance chain between the data and the dashboard is disrupted. Sarikaya et al. [1] emphasize the lack of provenance information, particularly within social organizational domains and urban informatics. In these contexts, data quality, trust, and accountability are of the utmost importance. Yet, current dashboards frequently fall short in eliciting these aspects, despite their pivotal role in enhancing public trust. Matheus et al. [17] conducted a study to understand and facilitate the design of dashboards that promote transparency and accountability. The research by Ivankovic et al. [18], which analyzed 158 COVID-19 dashboards, found that 34 (24.7%) did not disclose their data sources. Jarske et al. [19] proposed a framework to facilitate the effective communication of practical data provenance and to provide direction for forthcoming advancements in provenance visualization techniques tailored specifically for dashboards.

Concerns about the reliability of the data available on the Web are almost as old as the Web itself. The Web was invented in 1989; the first graphical Web browser, Mosaic, was created in 1993; and significant Web expansion began in 1995 [20]. As early as 1997, the Web inventor, Tim Berners-Lee, foresaw reliability issues in Web applications and suggested that they should provide a button he called the “Oh, yeah?” button for users to press whenever they lost trust in the information provided. When pressed, the system would display a set of metadata and annotations that would allow the user to regain confidence [21].

Motivated by the lack of a standard for visualizing provenance metadata in dashboards and inspired by Tim Berners-Lee’s pioneering vision, this paper proposes a representation model to standardize, model, generate, capture, and visualize dashboard provenance.

The authors anticipate that a standardized model for dashboard provenance will be beneficial, assisting architects and developers in building more effective, user-friendly provenance visualization interfaces and enabling dashboard end-users to access comprehensive, reliable provenance information about the data and visualizations presented on the dashboard. This improved understanding should empower users to trust the data and make well-informed decisions based on the provenance information.

By adopting a standardized model, the visualization of provenance should become

more consistent and coherent, making it easier for users to comprehend and interpret the origins and transformations of the displayed data and visualizations. A standardized model encourages interoperability and uniformity across dashboard applications, enhancing collaboration and making it easier to share provenance data across platforms.

The need for a Dashboard Provenance Model

Just as different knowledge domains have specific applications and a need for provenance, dashboards also have their specificities. In this section, we present a dashboard provenance model.

The proposed model extends the W3C-PROV model to address the particular needs of dashboards, taking into account the common entities, collections of entities, activities, and agents involved in the typical lifecycle of a dashboard from its demand, construction, operation, and maintenance. This was identified during the Infovis for Public Health Research Project and in the study by Ivankovic et al. [18], which examined the COVID-19 dashboards chosen for the study.

The W3C-PROV offers a generic provenance data model that can be tailored and implemented in a wide range of domains, demonstrated by the 38 use cases proposed by the Provenance Incubator Group[22]. Other examples of applications include healthcare exchange [23], educational resources [24], geoprocessing workflows [14], and internet art [9]. Although each dashboard developer can use W3C-PROV to create their own solutions for displaying provenance information to their users, our research focuses on providing a standardized model that can be easily applied to any dashboard.

A standardized model for dashboard provenance should specify the relationships and hierarchies among dashboard entities (both visual and data), outline the customary activities involved, and delineate the roles typically undertaken by agents throughout a dashboard's lifecycle.

As stated by Ragan et al. [25], there exist five different types of provenance information: i) data provenance: encompasses the lineage of alterations and movements of data, which may encompass activities like data subsetting, merging, formatting, transformations, and simulations aimed at generating or incorporating new data; ii) visualization provenance: pertains to the sequence of graphical views and states of visualizations; iii) interaction provenance: involves a record of user actions and commands executed within a system; iv) insight provenance: captures the progression of cognitive outcomes and information attained from the analysis process, including analytical findings, hypotheses, and deductions; and v) rationale provenance: encompasses the chronicle of reasoning and intentions behind decisions, hypotheses, and interactions. This work will focus

on data and visualization provenance, including information about data sources, preprocessing steps, and design choices made during the creation of the visual element. However, it is important to highlight that this study will not focus on the visualization provenance resulting from user interaction with the dashboards, including the record of visualization states.

The construction of a dashboard provenance model aims to provide a consistent, comprehensive representation of the origins and histories of the components that make up dashboards, enabling the creation of provenance visualizations that facilitate a clear, accurate understanding of the context in which a particular dashboard was created.

In the early days of W3C-PROV development, the Provenance Incubator Group provided a list of provenance dimensions to guide the design of use case scenarios [13]. The provenance dimensions were grouped into content, management, and use. The content of provenance determines its meaning in a given context, and the modeling of provenance identifies its basic elements (entities, agents, and activities), relationships and properties among these elements, evolution and versioning, justifications for decisions, and expected results. The management of provenance leads to issues related to publication (visualization), access (browsing and querying), dissemination control (access control, licensing, and legal issues), and scale (volume) of provenance content. Finally, the use of provenance leads to issues with the user experience of the provenance visualization.

The design of use-case scenarios for dashboard provenance can vary depending on the dashboards' goals and user-specific needs. Using a provenance model for dashboards can improve the design of provenance, ensuring it aligns with the provenance dimensions and supports achieving user needs and the dashboard's goals.

The proposed model for dashboard provenance aims to provide metadata that enables users to evaluate the reliability, quality, and consistency of the content displayed on dashboards that include both visual and data artifacts.

A dashboard can be interpreted as a collection of visual entities and data, agents, and activities. Visual entities can include charts, maps, tables, infographics, or text. Data entities can be provided in different formats, such as CSV (Comma-Separated Values) [26], JSON (JavaScript Object Notation) [27], GEOJSON (JavaScript Object Notation-based geospatial data interchange format) [28], among others commonly used on the Web. It is necessary to map all individual entities and those that are part of a visual and data artifact collection that make up the dashboard for a comprehensive representation of provenance. Agents can be organizations, software systems, or individuals responsible for creating, maintaining, and updating the dashboard. Activities represent the actions or operations performed by agents

within the dashboard's lifecycle, such as creating, modifying, or updating visual components, data entities, or the dashboard itself.

The agents responsible for the dashboard can be categorized as follows:

- Demand-side agents: these are the agents (organization, department, or people) who initiated the development of the dashboard. They represent the stakeholders who identified the need for the dashboard to support their specific requirements and goals.
- Development and maintenance agents: these agents are responsible for creating, designing, and maintaining the dashboard's visual elements and data. They can be individuals or teams within the organization, or external contractors, who handle the technical aspects of dashboard development and ensure its operation.
- Curation agents: responsible for curating and managing the visual and data entities displayed on the dashboard. They ensure quality, accuracy, and reliability;
- Data update agents: these agents are responsible for regularly updating the data presented on the dashboard. They can be data scientists, analysts, database administrators, or automated processes that retrieve new data, perform data transformations, and incorporate updated information into the dashboard.

The activities involved in creating, maintaining, and updating the dashboard include the following:

- Creation and maintenance: involves the design, development, and maintenance of the dashboard, including the selection of visual elements, data sources, and overall layout. It encompasses tasks such as collecting and organizing data, creating visualizations, and integrating various components into a cohesive dashboard.
- Sustainability: ongoing efforts and practices to maintain and preserve the stability, viability, and longevity of the dashboard, visual entities, and data. Sustainability activities include regularly monitoring and evaluating the dashboard's performance, gathering user feedback and incorporating it into future updates, addressing technical issues and ensuring compatibility with new technologies or data sources, and aligning the dashboard with evolving user needs and organizational goals.
- Data update: Data update activity refers to updating the data used in a dashboard. It involves retrieving new data, incorporating it into the existing dataset, and ensuring the dashboard reflects the latest information.

By documenting the above activities, it becomes possible to understand the dashboard's development and evolution over time.

To identify basic provenance related to visual elements, it is proposed to divide the dashboard into three visual layers: i) the dashboard as a whole; ii) the dashboard subtopics;

and iii) individual visual entities. Data entities can be associated with one or more of these visual layers.

The division into layers aims to provide users with an understanding of provenance at different levels of detail, allowing the display of provenance information for individual visual entities, their relationships with other visual entities grouped under a common theme, and the dashboard as a whole. This approach enables users to access different levels of provenance, depending on their interests, ranging from more detailed to consolidated information.

Figure 2 shows the first layer of provenance, encompassing the dashboard as a whole. At this level, the dashboard is considered a single entity and a collection of other entities. This layer serves to provide an overview of the dashboard and achieve the following objectives: i) describe the general properties of the entity, such as title, descriptive text, global references, annotations, version, and creation date; ii) present the agents responsible for the dashboard curation, creation, and maintenance, including name, uniform resource locator (URL), description, and annotations; iii) capture the activities involved in the creation and maintenance of the dashboard.

Figure 3 shows the second layer of provenance, which corresponds to the subtopics addressed in the dashboard. For example, in a COVID-19 dashboard, distinct topics such as case and death counts, vaccination, and associated comorbidities can be addressed. Similarly to the first layer of provenance, in the second layer, entities can be understood as unique entities or as a collection of entities related to a specific subtopic addressed within the dashboard. Each subtopic may involve different entities, agents, and activities, for which maintaining and displaying provenance metadata is important. This allows the recording of properties and provenance information related to the addressed subtopics, the agents responsible for the curation, creation, maintenance, and updating of the second layer, and the activities performed.

For some dashboards, the definition of a second layer of provenance may not apply, as there may be no distinguishable subtopics that warrant this division.

Finally, Figure 4 depicts the third layer of provenance, referring to individual entities such as graphs, tables, or explanatory text. As with collections of entities, maintaining provenance metadata, agents, and performed activities is also necessary for individual entities.

ISSN: 2177-3246

Figure 2

Dashboard provenance - first layer

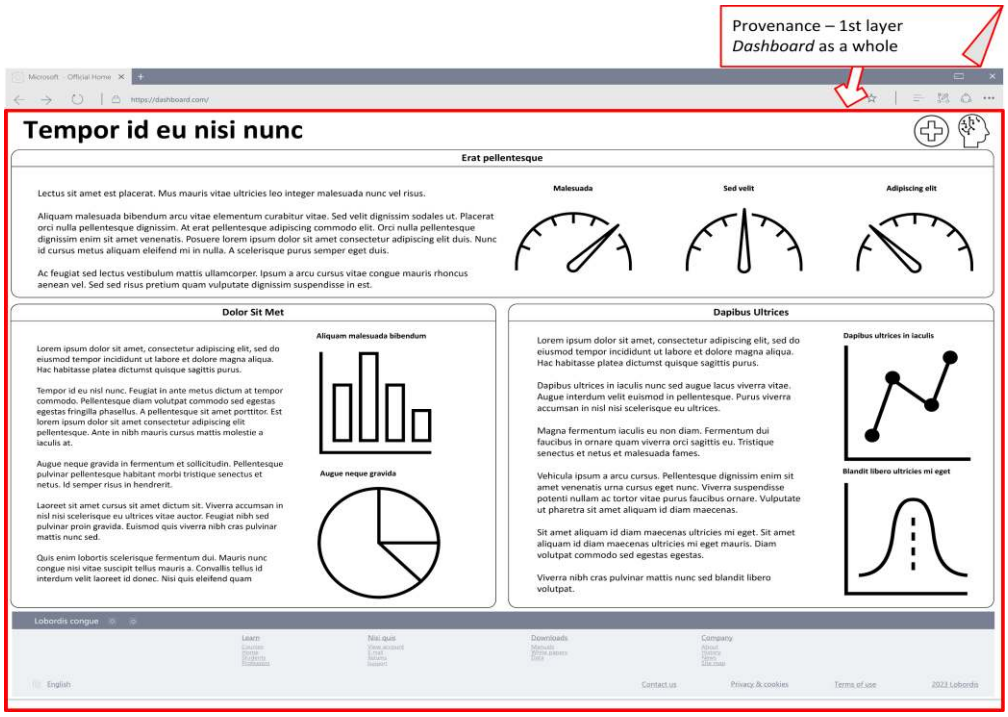
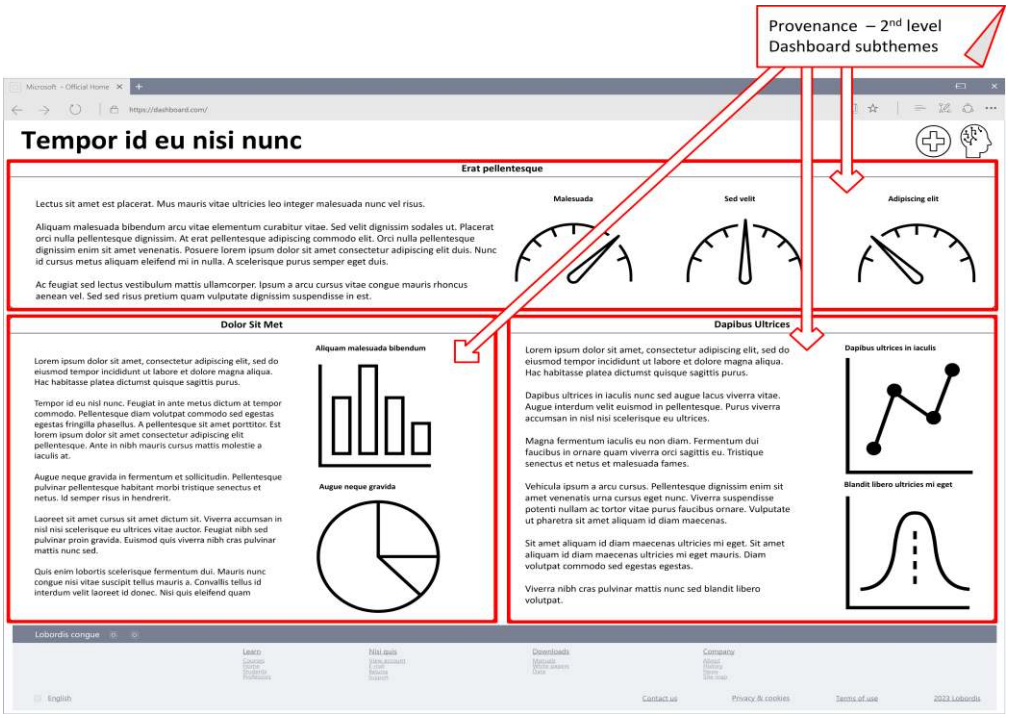


Figure 3

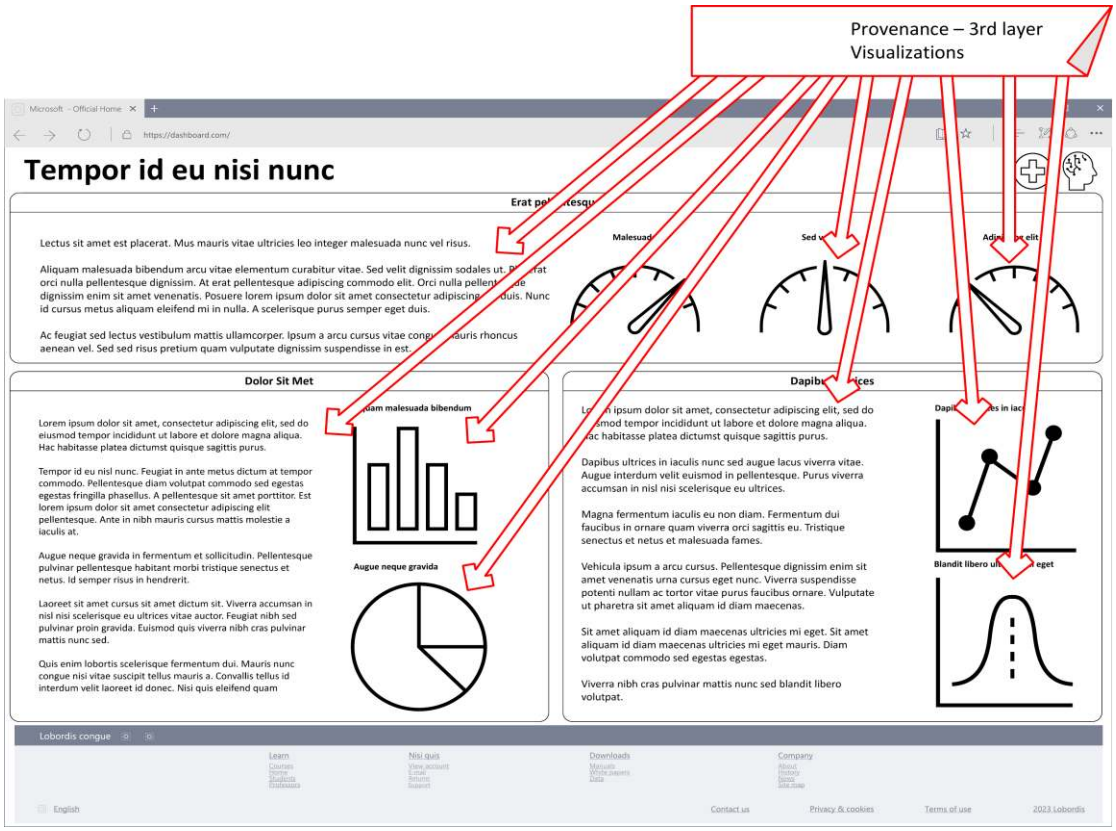
Dashboard provenance - second layer



Dashboard provenance - third layer.

Figure 4

Dashboard provenance – third layer



Figures 2–4 illustrate how the proposed dashboard's provenance can be structured in layers, providing the user with varying levels of information.

Regardless of the layer in which the basic elements of dashboard provenance are located, the following properties can be included:

- Entities:

– Visual: name, identification, description, creation date, versioning, authorship, maintainer, ownership, source, data sources, metadata source, relationships, annotations, and explanations (that should include validity and quality measures, generation parameters, graph construction steps, how the data attributes are mapped to the graph elements, graph interactions, data associated with individual nodes and edges in the graph).

– Data: name, identification, type, creation/maintenance date, frequency of update, version, authorship, ownership, maintainer, responsibility, source, metadata source, relationships, annotations, and explanations (which should include data collection methods, data processing steps, and data transformations).

- Activities: name, identification, type, description, inputs, outputs, parameters, dependencies, start time, end time, frequency of execution, authorship, relationships, versioning, validation, and quality measures.
- Agents: name, identification, role, contact information, description, responsibilities, affiliation, contributions, relationships, trustworthiness.

Table 1

Set of metadata for dashboard provenance

Metadata	Description	Application
prov:agent	Agents are the organizations, departments, and sectors responsible for the dashboard.	To identify the agents involved in the development of the dashboard, as well as those responsible for the maintenance of visual entities and data, data and visual artifacts curation, and data updating activities.
prov:entity	Entities can include the dashboard itself, its visual components, groups of visual components, and the data.	To identify the visual and data entities. In dashboards, visual entities are derived from data entities or other visual entities.
prov:actity	On the dashboards, there are activities for creating and maintaining visual and data entities, as well as for data updates.	The activities of development and maintenance of visual and data entities are important to understand when the entities were created and modified due to evolutionary or corrective maintenance. There may be a single activity for data updates or specific activities for each dataset in the dashboard.

Table 2

W3C-PROV - basic elements

Metadata	Description	Application
prov:wasGeneratedBy	Specifies which activity generated the entity.	Connects a specific entity to the activity that generated or transformed it. This can be an activity that involves developing or maintaining a visual entity or updating a dataset.
prov:wasDerivedFrom	Connects a visual entity to other visual or data entities that are its source or origin.	Inform the user which entities a given entity was derived from.
prov:wasAttributedTo	Specifies which agent a given entity is assigned.	Inform the user to which agent a given entity has been assigned.
prov:startedAtTime	Specifies when the activity was started.	Notifies the user if a given activity was started.
prov:endedAtTime	Specifies when a given activity was ended.	Inform the user if a given activity has ended.
prov:used	Specifies whether an existing entity was modified or a given activity generated a new one, and the entity used. For example, in a dashboard, it can indicate that a specific dataset was used to update the dashboard data or that a new version of an entity was produced using an existing one.	Informs the user which entity was used to modify an existing entity or generate a new entity.
prov:wasAssociatedWith	Specifies the agent responsible for executing an activity.	Inform the user which agent (software or organization) is responsible for updating a dataset or maintaining a specific visual component.
prov:actedOnBehalfOf	Specifies hierarchical relationships or delegated responsibility between organizations and departments.	Displays hierarchical or delegated responsibility among agents, giving users an

Metadata	Description	Application
		understanding of the relationships among them.

Table 3

W3C-PROV - basic relations

Metadata	Description	Application
prov:generatedAtTime	Specifies the time at which the entity was generated.	Informs the user about the entity's generation time. It can indicate, for example, the date and time when a particular dataset was last updated. This information clarifies the data's freshness and accuracy within the entity, enabling users to make informed decisions based on the latest available data.
prov:collection	Specifies an entity composed of a set of visual and data entities.	The statement that an entity consists of a set of entities is valuable for establishing relationships between them, because it means that they share common data sets or related information.
prov:softwareAgent	Specifies that a given agent is a software component.	This property explicitly states that the entity responsible for executing an activity is a software component. For example, it can indicate that a software program performs daily updates on a dataset.
prov:organization	Specifies that a given agent is an organization.	It can be used to indicate that a specific agent responsible for the dashboard or dataset is an organization.
prov:hadPrimarySource	Specifies that a given entity was derived from a primary source.	Informs the user that a specific entity has a primary data source. Help users understand the origin and reliability of the data presented in the entity.
prov:type	Specifies if a given agent is an organization, department, person, or software.	The user is informed about the type of agent involved in the activity or responsible for the entity.

Table 4

W3C-PROV - extended relations

Metadata	Description	Application
dash:name	Name assigned to the provenance element (entity, agent, or activity), which will be visible to the user.	The user can identify the provenance element by their name, which could be the same as the title of a graph (visual entity) or the name of the dashboard publisher (agent).
dash:version	The identification mechanism is particularly significant when there are multiple iterations or updates throughout the dashboard lifecycle. A version number provides traceability, transparency, compatibility, and facilitates effective communication regarding different dashboard versions.	
dash:role	Describing the role or function of the agent in the overall process, such as demand-side agent, development and maintenance agent, data curation agent, and data update agent. By understanding the role of an agent, the user can better grasp the agent's responsibilities in the dashboard	

Metadata	Description	Application
	context, thereby conveying credibility and transparency.	
dash:description	A textual description that provides additional information about the provenance element (entity, agent, or activity), including its purpose, content, or context.	The additional information provided to the user in a dashboard may contextualize the topic addressed, providing clarity and understanding.
dash:annotations	Relevant comments or observations associated with the provenance element, providing additional details or clarification on its nature or content. This includes defining technical terms and specific concepts, explaining measurement units, and describing the formulas and calculations used, among other tasks. Annotations can be interpreted as extended descriptions.	This information helps the user correctly interpret visualizations and data, avoiding misunderstandings. Annotations could also supply information about agents (rules, responsibilities, contributions, and contact information) and activities (inputs, outputs, parameters, dependencies, validation, and quality measures)
dash:contactInformation	Providing contact details, such as email address, phone number, or organizational affiliation	Contact information allows users to reach out to the agent for further information or inquiries.
dash:trustworthiness	Any certification, credential, or trust metrics associated with the agent.	This information assures users of the agent's reliability and expertise in the field.
dash:url	The URL refers to the Web address where users can access additional details or resources related to the entities, agents, and activities in the dashboard.	By including URLs that provide information about entities, agents, and activities, users can access and verify the sources and background information for the provenance resources displayed in the dashboard, enhancing their understanding and trust in the data and visualizations presented.

Source: provided by the authors, based on the W3C-PROV specifications.

Table 1 presents a compilation of provenance metadata as defined by W3C-PROV [15], along with additional properties tailored for dashboard provenance. Table 2 provides a detailed exposition of the fundamental constituents within W3C-PROV. Furthermore, Tables 3 and 4 outline the elementary and expanded relations of W3C-PROV that are expected to be employed within the realm of dashboard provenance. Table 5 presents a collection of custom metadata under the **dash** prefix that extends the W3C-PROV framework to meet the requirements of dashboard provenance. This extension enriches the W3C-PROV with additional context and provenance details.

Applying the dashboard provenance model

This modeling process sample uses a preexisting dashboard, although it is highly recommended to define the dashboard's provenance in parallel with its design.

ISSN: 2177-3246

Figure 5

Dashboard COVID-19 no Brasil - first layer

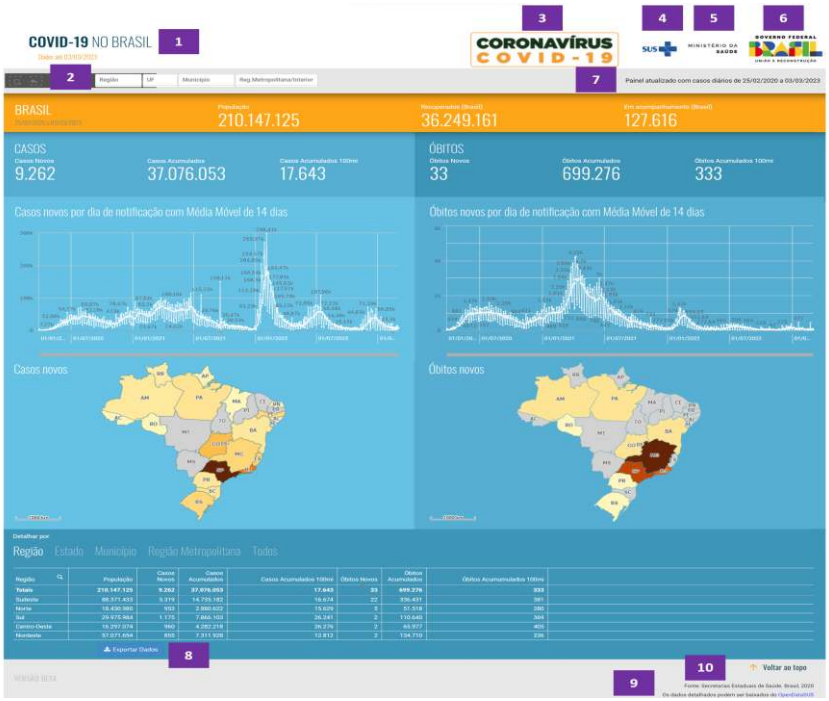


Figure 6

Dashboard COVID-19 no Brasil - second layer

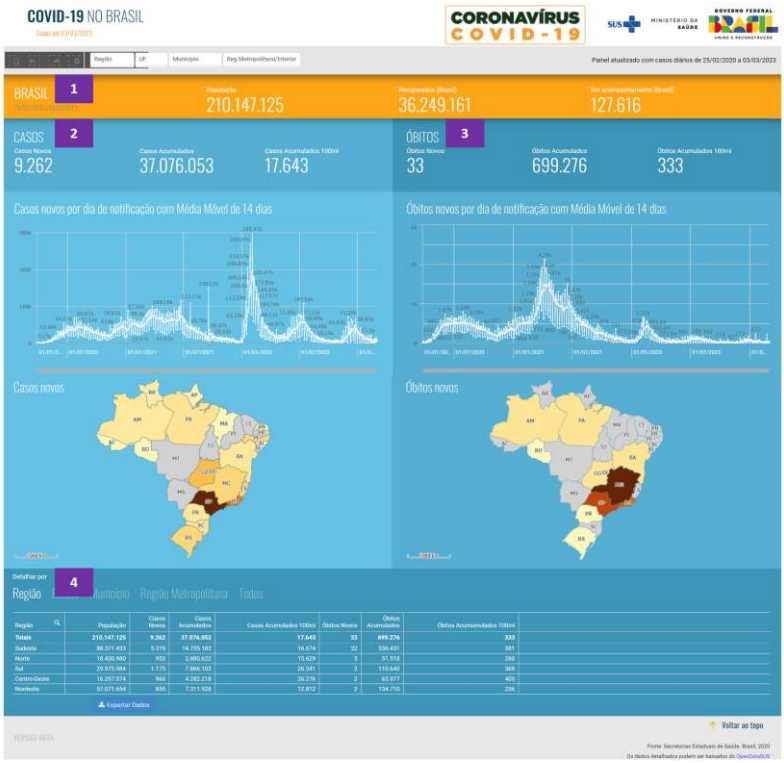
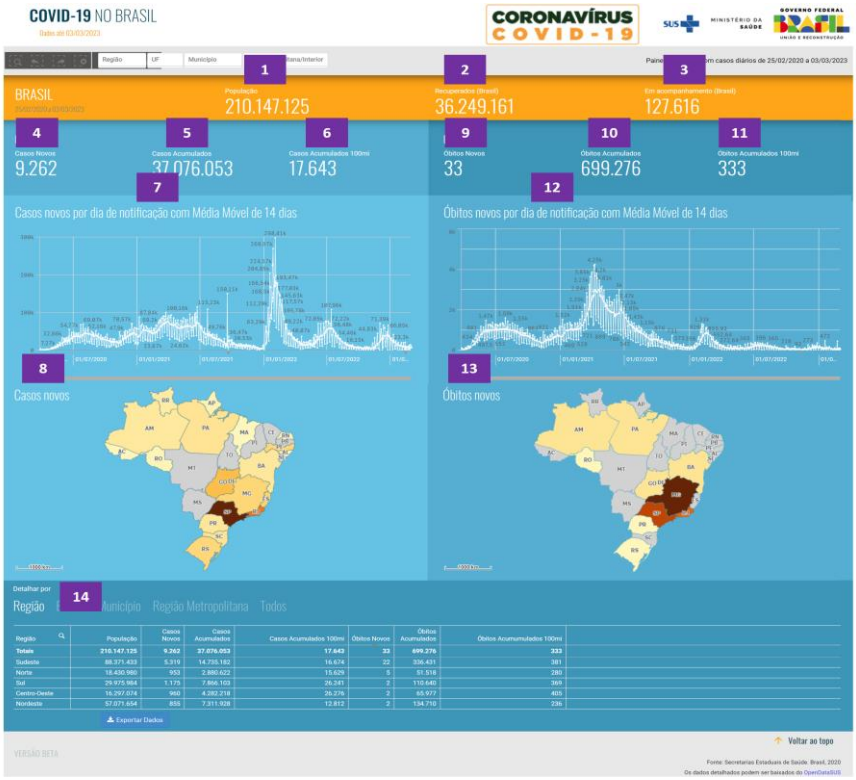


Figure 7

Dashboard COVID-19 no Brasil - third layer



Figures 5–7 display the hierarchical structure of the three proposed layers of dashboard provenance employed in the COVID-19 no Brasil context [29].

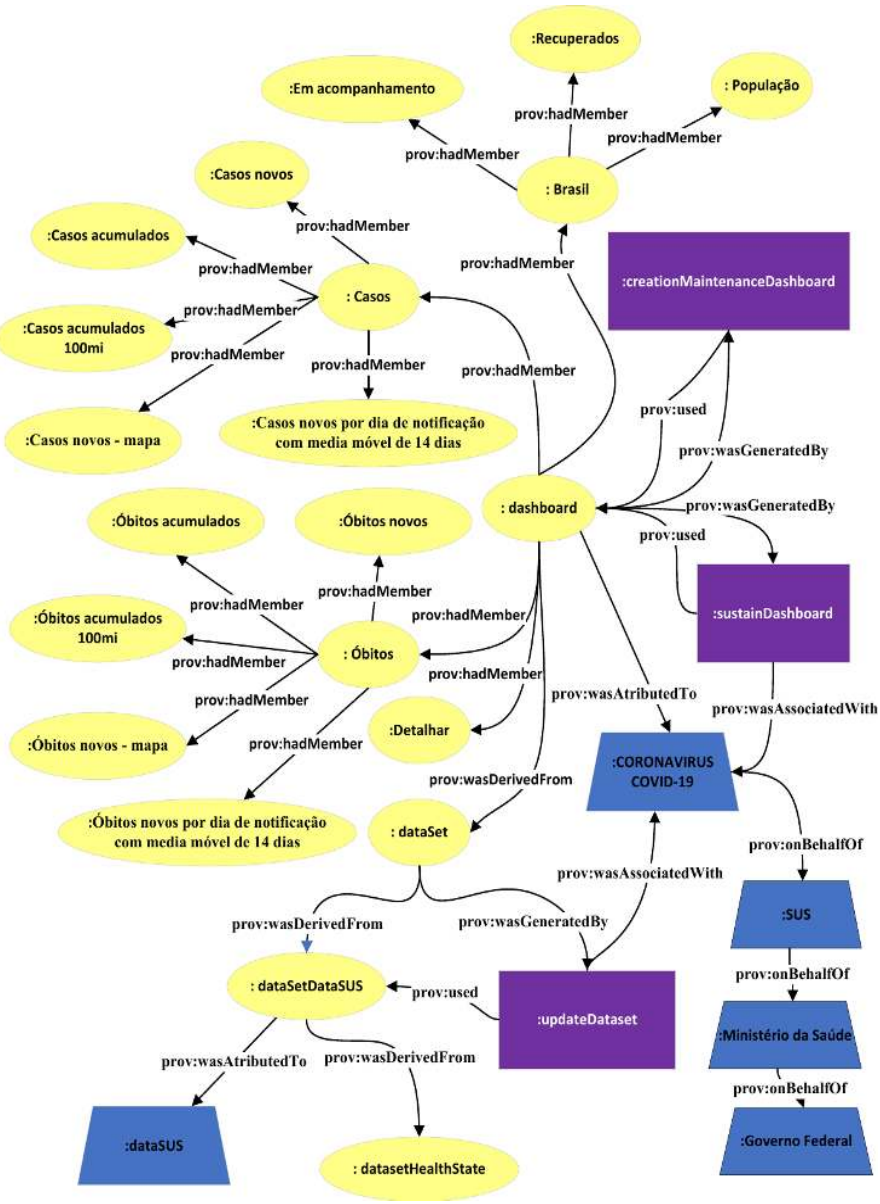
On the first level of provenance, the dashboard as a whole, shown in Figure 5, a set of related provenance metatada was identified: 1) the name of the dashboard; 2) the last date covered by the dashboard (data until 2023-03-03); 3) the logo used by the Brazilian government to represent the COVID-19 taskforce; 4) the logo of the Unified Health System (SUS - *Sistema Único de Saúde*); 5) the logo of the Brazilian Ministry of Health (*Ministério da Saúde*); 6) the logo of the Brazilian Federal Government (*Governo Federal do Brasil*); 7) the interval of data displayed on the dashboard - Dashboard updated with daily cases from 2020-02-25 to 2023-03-03; 8) download the data; 9) the data source - Brazilian State Departments of Health (*Secretarias Estaduais de Saúde. Brasil*) - collection of agents; and 10) the curator of the data, DataSUS (IT department of the Brazilian Unified Health System) is the agent responsible for data curation. Items 3, 4, 5, and 6 represent a hierarchy of agents: 3 acted on behalf of 4, which acted on behalf of 5, which acted on behalf of 6.

On the second level of provenance, shown in Figure 6, the following visual subthemes were identified: 1) *Brasil* - general data about COVID-19 in Brazil; 2) *Casos* - number of infected people; 3) Óbitos - number of deaths; and 4) *Detalhar por* - detailed number of cases.

On the third level of provenance, shown in Figure 7, the following visual entities were identified: 1) *População* - Brazilian population card; 2) *Recuperados (Brasil)* - Brazilian number of recovered cases card; 3) *Em acompanhamento (Brasil)* - Brazilian number of monitored cases; 4) *Casos novos* - new cases; 5) *Casos acumulados* - number of accumulated cases; 6) *Casos acumulados 100mi* - number of accumulated cases per 100,000 inhabitants; 7) *Casos novos por dia de notificação com media móvel de 14 dias* - new cases per day of notification with 14-day moving average graph; 8) *Casos novos* - new cases on a Brazilian choroplet map; 9) *Óbitos novos* - new cases of deaths; and 10) *Óbitos acumulados* - number of accumulated cases of deaths;

Figure 8

Provenance resources of the COVID-19 Brasil dashboard



11) *Casos acumulados 100mi* - number of accumulated cases of death per 100,000 inhabitants; 12) *Óbitos novos por dia de notificação com media móvel de 14 dias* - new cases of death per day of notification with 14-day moving average graph; 13) *Óbitos novos* - new cases of death on a Brazilian choroplet map; 14) *Detalhar por* - Table showing COVID-19 indicators in Brazil at different aggregation levels.

The *COVID-19 no Brasil* dashboard [29] is depicted in Figure 8, which shows the dashboard's basic provenance structure, including its elements (activities, entities, and agents) and basic and extended relationships.

The following sample of provenance ontology exemplifies how the proposed dashboard provenance model extends the W3C-PROV and provides metadata to describe the provenance of the COVID-19 Brasil dashboard [29].

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dash: <http://dash.com/dash#> .
@prefix prov: <http://www.w3.org/ns/prov#> .

:dashboard
a prov:Entity, prov:Collection;
dash:Name "COVID-19 NO BRASIL";
dash:Version Beta
dash:Description “Painel atualizado com casos diários de 25/02/2020 a 03/03/2023”
dash:annotations “Em atendimento à definição clínica oficial da condição pós-Covid-19 pela Organização Mundial da Saúde (OMS).”
dash:url “https://infoms.saude.gov.br/extensions/covid-19.html”
prov:wasGeneratedBy :creationMaintenanceDashboard
prov:wasDerivedFrom :dataset;
prov:wasAttributedTo :CORONAVIRUS COVID-19;
prov:hadMember: “Brasil”, “Casos”, “Óbitos”, “Brasil”, “Detalhar”;
.
:Casos
a prov:Entity, prov:Collection;
dash:Name "Casos";
dash: “1.0.0.0”
dash:Description “Casos novos por dia, por dia de notificação.”
Os casos novos são calculados como a diferença entre os casos informados da semana

atual e os casos acumulados da semana epidemiológica anterior."
dash:url "https://infoms.saude.gov.br/extensions/covid-19.html"
prov:wasGeneratedBy :creationMaintenanceDashboard
prov:wasDerivedFrom :dataset;
prov:wasAttributedTo :CORONAVIRUS COVID-19;
prov:hadMember: "Casos novos por dia, por dia de notificação com média móvel de 14 dias", "Casos novos",
.
:CORONAVIRUS COVID-19 #COVID-19 taskforce
a prov:Agent;
dash:Name "CORONAVIRUS COVID-19 taskforce";
dash:Role "Developer", "Maintenance"
dash:Description "Força-tarefa criada para acompanhar a evolução da pandemia do COVID-19 no Brasil."
dash:trustworthiness "gov.br."
dash:url "https://saude.gov.br/covid-19/covid-19-taskforce.html"
prov:actedOnBehalfOf :SUS;
.
:SUS
a prov:Agent;
dash:Name "Sistema Único de Saúde"; dash:Role "demanding";
dash:Description "Sistema Único de Saúde é a denominação do sistema público de saúde brasileiro criado pela Constituição Federal de 1988."
Em todo o país, o SUS deve ter a mesma doutrina e a mesma forma de organização, sendo definido, na Constituição, como um conjunto de elementos doutrinários e organizacionais do sistema de saúde: os princípios da universalização, da equidade, da integralidade, da descentralização e da participação popular."
dash:contactInformation "Esplanada dos Ministérios - Bloco G, Edifício Sede - CEP: 70058-900 - Brasília/DF"
dash:trustworthiness "gov.br"
dash:url "https://www.gov.br/saude/pt-br"

2 DISCUSSION AND FUTURE WORK

This paper addresses the need for a dashboard provenance model to provide comprehensive information about the visual and data entities within a dashboard.

Additionally, we have emphasized the importance of identifying the agents involved in the design, development, and maintenance of the dashboard, as well as those responsible for tasks such as data curation and ongoing support to ensure its sustainability.

A key contribution of this work is the development of a dashboard provenance model. This model provides a structured approach to capturing and representing the provenance of dashboards, including the entities, agents, activities, and their properties and relationships. The next steps of this ongoing project must include:

- **Development of a formal data model for dashboard provenance**: After defining the main components of the provenance of the dashboard, it is important to define a data model that represents the structure and organization of the provenance data associated with the dashboard. Based on the W3C-PROV-DM [30], the data model should specify how different elements of provenance, such as agents, activities, entities, their relationships, and properties, are represented and interconnected within the dashboard. The data model should provide a comprehensive and standardized framework for storing, querying, and analyzing provenance data. As this study focuses on Web-based dashboards, the PROV-JSON format [31], [32], [33] should be considered.
- **Verification and validation of the model**: To determine how reliable the data generated by the model is, a verification and validation process will be conducted. In conjunction with the analysis of the dashboards extracted from the study of Ivankovic et al. [18], the expertise acquired in the analysis and production of different dashboard prototypes, and the dashboard design system at Infovis Public Health Research Project. Commencing with the guidelines delineated within the W3C-PROV standard and recent advancements in provenance research, this evaluation should encompass an appraisal of the credibility of the proposed metadata constructs. Furthermore, it should scrutinize the model's potential for generalization across diverse contexts and its feasibility for replication, ensuring the reliability and reproducibility. The prototypes will be created for assessment by recognized experts in the field to obtain their professional evaluation of the model.
- **Design a user interface for visualizing the dashboard provenance**: Analyzing provenance metadata is not a common task for the general public, particularly dashboard users, and requires literacy in provenance and provenance visualization [34]. Users may become overwhelmed in trying to understand and explore the meaning of data provenance due to its large volume and complexity. Data provenance visualization should be seamlessly integrated into the dashboard to enable

satisfactory understanding. The use of infographics and storytelling approaches has been little explored in the literature of representing provenance. Still, it can be an appropriate solution for helping dashboard users to visualize provenance.

- **Assessment of the user's experience in navigating the dashboard's provenance**: The assessment of the user's experience with the visualization of the dashboard provenance involves assessing how well the provenance information is presented and utilized within the dashboard interface, to understand how users will perceive and interact with the provenance features and whether they will enhance their understanding and trust in the displayed information due to provenance. The evaluation should include user interviews, surveys, and user experience tests. The key aspects to consider during the evaluation are: user's understanding of the provenance information and its relevance to the dashboard content, the usability and accessibility of the provenance features within the dashboard interface, the impact on decision making, user overall satisfaction with the presentation of provenance information, and detect user preferences, suggestions for improvements, and whether they find the provenance features valuable and beneficial. This feedback can guide improvements to enhance the user experience and ensure effective use of provenance in dashboards.

3 CONCLUSION

This paper highlights the importance of incorporating provenance into dashboards to improve decision-making. Provenance provides essential information about the dashboard's origin and the processes involved in its design, development, maintenance, and sustainability. It also identifies the various actors, such as institutions, departments, and teams, involved in the dashboard's lifecycle.

In contrast to traditional approaches that focus on representing data at its final processed stage, provenance-driven dashboards enable users to see a more nuanced perspective on the curatorial history of the data and the visualizations presented in the dashboard. By including provenance in dashboards, users can make more informed decisions based on a comprehensive understanding of the dashboard's context and reliability.

We have presented a model that aims to identify, capture, and organize the entities, agents, and activities, along with their attributes, properties, and relationships, to create a comprehensive model of dashboard provenance.

We recognize that significant effort will be required to create a comprehensive model

of dashboard provenance. This is an ongoing endeavor, and further progress is needed to finalize the data provenance model and enable the design of standardized interfaces for displaying dashboard provenance.

The model also requires thorough evaluation regarding its comprehensiveness and validity. This entails assessing the soundness of each suggested metadata construct, examining whether the model can be applied universally, and verifying that it can be replicated to yield consistent outcomes.

We hope this paper will spur further development of dashboards with stronger provenance information, thereby improving decision-making.

REFERENCES

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What do we talk about when we talk about dashboards? IEEE Transactions on Visualization and Computer Graphics, 25, 682–692. https://doi.org/10.1109/TVCG.2018.2864903

Shankar, K., Jeng, W., Thomer, A., Weber, N., & Yoon, A. (2021). Data curation as collective action during COVID-19. Journal of the Association for Information Science and Technology, 72, 280–284. https://doi.org/10.1002/ASI.24406

Moreau, L., & Groth, P. (2013). Provenance: An introduction to PROV (Vol. 3, pp. 1–131). https://doi.org/10.2200/S00528ED1V01Y201308WBE007

Centers for Disease Control and Prevention. (2022). COVID data tracker. https://covid.cdc.gov/covid-data-tracker

Johns Hopkins University. (2022). Coronavirus resource center. https://coronavirus.jhu.edu/

Florida Department of Health. (2022). Florida COVID action. https://floridacovidaction.com/

Tompkins, A., & Humphries, L. (2020). Provenance research today: Principles, practice, problems (1st ed.). Lund Humphries.

Fuhrmeister, C., & Hopp, M. (2019). Rethinking provenance research. Getty Research Journal, 11(1), 213–231. https://doi.org/10.1086/702755

Rossenova, L., de Wild, K., & Espenschied, D. (2019). Provenance for internet art using the W3C PROV data model. In Proceedings of the 16th International Conference on Digital Preservation (iPRES 2019) (pp. 16–20).

Gramlich, J. (2017). Reflections on provenance research: Values – politics – art markets. Journal for Art Market Studies, 1(2), 322–327. https://doi.org/10.23690/jams.v1i2.15

Cheney, J., Chong, S., Foster, N., Seltzer, M., & Vansummeren, S. (2009). Provenance: A future history. In Proceedings of OOPSLA (pp. 957–964). https://doi.org/10.1145/1639950.1640064

Moreau, L., & Foster, I. (2006). Provenance and annotation of data: International Provenance and Annotation Workshop (IPAW 2006) (Vol. 1). Springer.

Provenance Incubator Group. (2010). Provenance dimensions. https://www.w3.org/2005/Incubator/prov/wiki/Provenance_Dimensions

Zhang, M., Jiang, L., Zhao, J., Yue, P., & Zhang, X. (2020). Coupling OGC WPS and W3C PROV for provenance-aware geoprocessing workflows. Computers & Geosciences, 138, 104419. https://doi.org/10.1016/j.cageo.2020.104419

Groth, P., & Moreau, L. (2013). PROV-overview – an overview of the PROV family of documents. W3C Working Group. https://www.w3.org/2005/Incubator/prov/wiki/Use_Case_Template

Zhang, Y., Sun, Y., Gaggiano, J. D., Kumar, N., Andris, C., & Parker, A. G. (2022). Visualization design practices in a crisis: Behind the scenes with COVID-19 dashboard creators. https://doi.org/10.48550/arXiv.2207.12829

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. Government Information Quarterly, 37, 101284. https://doi.org/10.1016/J.GIQ.2018.01.006

Ivanković, D., et al. (2021). Features constituting actionable COVID-19 dashboards. Journal of Medical Internet Research, 23. https://doi.org/10.2196/25682

Jarske, J. M., Rady, J., Filgueiras, L. V. L., Velloso, L. M., & Santos, T. L. (2022). Data provenance visualization in Brazilian public health dashboards. TREX 2022 Workshop on Trust and Expertise in Visualization.

Brügger, N., & Milligan, I. (2018). The SAGE handbook of web history (1st ed.). SAGE Publications.

Moreau, L. (2010). The foundations for provenance on the web. Foundations and Trends in Web Science, 2, 99–241. https://doi.org/10.1561/1800000010

HL7 Community. (2022). HL7 FHIR specification: Resource provenance – content. https://www.hl7.org/fhir/provenance.html

dos Santos, R. R., Santos, M. T. P., & Ciferri, R. R. (2023). ProvOER model: A provenance model for open educational resources. Heliyon, 9, e13311. https://doi.org/10.1016/J.HELIYON.2023.E13311

Ragan, E. D., Endert, A., Sanyal, J., & Chen, J. (2016). Characterizing provenance in visualization and data analysis. IEEE Transactions on Visualization and Computer Graphics, 22, 31–40. https://doi.org/10.1109/TVCG.2015.2467551

RFC 4180. (2005). Common Format and MIME Type for CSV Files. https://www.rfc-editor.org/rfc/rfc4180

RFC 8259. (2017). The JavaScript Object Notation (JSON) data interchange format. https://tools.ietf.org/html/rfc8259

RFC 7946. (2016). The GeoJSON Format. https://tools.ietf.org/html/rfc7946

Ministério da Saúde. (2023). COVID-19 no Brasil. https://infoms.saude.gov.br/extensions/covid-19_html

Moreau, L., Missier, P., et al. (2013). PROV-DM: The PROV data model for provenance. https://www.w3.org/TR/prov-dm/

Werner, B., & Moreau, L. (2020). ProvViz: An intuitive PROV editor and visualizer. In International Provenance and Annotation Workshop (pp. 231–236). Springer.

Zhai, J., Chen, H., & Yuan, C. (2017). Provenance metadata of open government data based on PROV-JSON. In ACM International Conference Proceedings Series (pp. 594–595). https://doi.org/10.1145/3085228.3085229

Huynh, T. D., Jewell, M. O., Keshavarz, A. S., Michaelides, D. T., Yang, H., & Moreau, L. (2013). The PROV-JSON serialization. https://www.w3.org/Submission/prov-json/#introduction

Yazici, I. M., Karabulut, E., & Aktas, M. S. (2018). A data provenance visualization approach. In 2018 14th International Conference on Semantics, Knowledge and Grids (SKG) (pp. 84–91). IEEE. https://doi.org/10.1109/SKG.2018.00019